

Research Article

Distributed 3D Source Localization from 2D DOA Measurements Using Multiple Linear Arrays

Antonio Canclini, Fabio Antonacci, Augusto Sarti, and Stefano Tubaro

Dipartimento di Elettronica Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milan, Italy

Correspondence should be addressed to Antonio Canclini; antonio.canclini@polimi.it

Received 17 May 2017; Accepted 11 September 2017; Published 25 October 2017

Academic Editor: Paolo Barsocchi

Copyright © 2017 Antonio Canclini et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This manuscript addresses the problem of 3D source localization from direction of arrivals (DOAs) in wireless acoustic sensor networks. In this context, multiple sensors measure the DOA of the source, and a central node combines the measurements to yield the source location estimate. Traditional approaches require 3D DOA measurements; that is, each sensor estimates the azimuth and elevation of the source by means of a microphone array, typically in a planar or spherical configuration. The proposed methodology aims at reducing the hardware and computational costs by combining measurements related to 2D DOAs estimated from linear arrays arbitrarily displaced in the 3D space. Each sensor measures the DOA in the plane containing the array and the source. Measurements are then translated into an equivalent planar geometry, in which a set of coplanar equivalent arrays observe the source preserving the original DOAs. This formulation is exploited to define a cost function, whose minimization leads to the source location estimation. An extensive simulation campaign validates the proposed approach and compares its accuracy with state-of-the-art methodologies.

1. Introduction

The problem of acoustic source localization with microphone arrays has been an active topic of research in the last thirty years [1]. Several are the applications that benefit from this technology, among which video-conferencing [2, 3], audio-surveillance [4], and so forth are worth mentioning.

In the literature, it has been proposed to use either compact or distributed arrays. Compact arrays are advantageous since their deployment is easy. As an example, in a video-conferencing system, a compact array can be located in the proximity of the camera to steer it towards the detected speaker. Unfortunately, compact arrays pose some limits in terms of localization accuracy. Indeed, for far sources, only the direction of arrival (DOA) can be accurately estimated, while the range estimate is generally unreliable [1]. On the other hand, distributed arrays envision the presence of multiple sensors, each equipped with one or more microphones, located around the volume of interest [5–10]. With this configuration, the source is observed from different angles, with benefits in terms of localization accuracy. Until recently,

the cost of the deployment of the sensors and the cabling made this kind of solutions cumbersome and possible only in specific ad hoc applications. In the last few years, however, the advent of wireless sensor networks made this kind of solutions attractive for a wider range of applications.

When each sensor is equipped with multiple microphones, the direction of arrival of the source can be measured internally to each sensor. The source is then localized by a central node, which combines measurements coming from the individual sensors. The advantage of DOAs with respect to other measurements (e.g., time differences of arrival or times of arrival) lies in the fact that each sensor transmits a single measurement to the central node, independently from the number of microphones.

Source localization from multiple DOAs can be brought to the problem of triangulating the estimated DOA lines [11, 12]. Following this approach, 3D localization is typically accomplished by combining multiple 3D DOAs measured at sensors with noncollinear microphone arrangements. For instance, multiple spherical arrays are used in [13]. A 3D DOA is represented by a pair of angles, which denote

the azimuth and the elevation of the source with respect to the sensor position. Therefore, measuring a 3D DOA involves a two-dimensional search space. As robust DOA estimation algorithms are generally based on grid-search solutions (e.g., beamforming [14], steered-response-power [15]), the inherent power requirement may be an issue for their implementation at local sensors. For this reason, many works limit source localization to 2D (planar) geometries, assuming that all the sensors and the source lie on the same plane. This is typically accomplished through the triangulation of multiple 2D DOAs, as done, for instance, in [8, 12, 16]. From a geometrical standpoint, a 2D DOA is measured in the plane containing the array and the source and consequently consists of a single angle. The search space is therefore one-dimensional, and simple array geometries can be adopted for this task. For instance, using a linear array, the DOA can be estimated by evaluating an objective function in the range $[-\pi/2, \pi/2]$, suitably sampled according to the desired angular resolution. Moreover, in order to achieve a prescribed resolution, a smaller number of microphones are required compared to that needed for 3D DOA measurements. For these reasons, 2D DOA measurements are more affordable for low-power sensing devices, as typically required in designing wireless sensor networks.

In this manuscript, we propose a technique that localizes a source in 3D from the combination of 2D DOAs, measured using multiple linear microphone arrays. Differently from state-of-the-art methods exploiting 2D DOAs, we remove the requirement of working in a planar geometry. Specifically, we do not pose any constraint about the positions of the source and of the arrays in the 3D space. The proposed technique proceeds in two steps. First, each array measures the 2D DOA of the source referred to a local plane, determined by the lying line of the array and the location of the source. Using the concept of *equivalent arrays*, the array positions are roto-translated into equivalent ones all lying in the same plane containing also the source. Each equivalent array position is found so that the original DOA is preserved, as well as its distance from the source. Once this transformation has been accomplished, in the second step the source can be localized in a 2D geometry. In order to do this, we take advantage of the Ray Space [17], in which DOAs are interpreted as acoustic rays originating from the source and impinging on the acoustic centers of the sensors. Acoustic rays are parameterized by the parameters of the line on which they lie. A cost function is defined, whose minimization gives us the source location.

The proposed technique is advantageous with respect to state-of-the-art methods for 3D source localization using DOAs, which, to the best of our knowledge, all rely on 3D measurements. Being based on simpler 2D DOA measurements, indeed, the proposed method requires a reduced number of microphones for each sensor of the network, with obvious advantages in terms of hardware cost (microphones and analog-to-digital converters). For the same reasons, power consumption at each sensor turns out to be significantly reduced, as the beamforming operation leading to the 2D DOA estimate is accomplished on a one-dimensional search space. Finally, it is worth noticing that, exploiting a simple

geometry such as that of a linear array, the manufacturing costs are much smaller than those required to realize more complex structures (e.g., spherical and cylindrical arrays).

The remainder of the manuscript is structured as follows. Section 2 gives some background information concerning the Ray Space representation and the localization of acoustic sources in 2D. Section 3 introduces the theoretical framework for the proposed approach, with particular reference to the equivalent arrays. Section 4 addresses the problem of source localization using the equivalent arrays. Section 5 is devoted to the validation of the proposed technique. Finally, Section 6 draws some conclusions.

2. Background

2.1. Ray Space Representation. Consider an acoustic source in a 2D scenario. Its Cartesian coordinates are (s_x, s_y) . In geometrical acoustics, it can be described as the set of all acoustic rays that originate from it. Each ray can be parameterized with the line on which it lies. The homogeneous representation of all the lines passing through (s_x, s_y) is

$$as_x + bs_y + c = 0. \quad (1)$$

This is equivalent to the vector form

$$\mathcal{J}^T \mathbf{l} = 0, \quad \mathcal{J} = [s_x, s_y, 1]^T, \quad \mathbf{l} = [a, b, c]^T. \quad (2)$$

The parameter vector \mathbf{l} is homogeneous; that is, any scaling $k\mathbf{l}$, $k \neq 0$, represents the same line. Consequently, the homogeneous coordinates \mathbf{l} form a class of equivalence in a two-dimensional projective space, defined in [18] as the *projective Ray Space*.

2.2. DOAs in the Ray Space. Let us now assume that the acoustic source is located in the far field with respect to the center of a microphone array (m_x, m_y) ; that is, the length of the array is much smaller than the curvature of the wavefront. In this context, the spatial information of the source can be described in terms of its direction of arrival (DOA), that is, the angle ϑ of propagation of the wavefront. Source and array location are related to the DOA by

$$\vartheta = \arctan \frac{s_y - m_y}{s_x - m_x}. \quad (3)$$

The DOA defines the acoustic ray that joins the source position and the observation point, which turns to be oriented as ϑ . In the Ray Space, the DOA is thus represented as the acoustic ray passing through (m_x, m_y) and directed as ϑ , whose parameters are

$$\begin{aligned} a &= \sin \vartheta, \\ b &= -\cos \vartheta, \\ c &= m_x \sin \vartheta - m_y \cos \vartheta. \end{aligned} \quad (4)$$

2.3. 2D Source Localization in the Ray Space. We now consider the problem of localizing the acoustic source, by

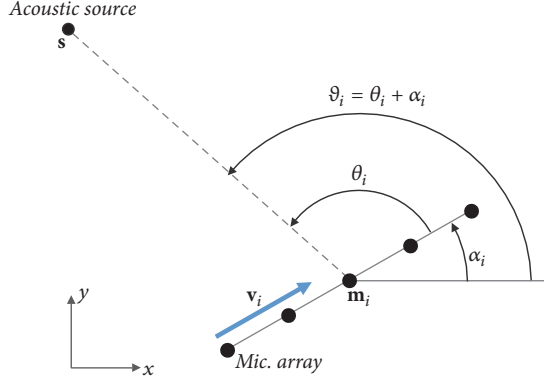


FIGURE 1: A microphone array centered at $\mathbf{m}_i = [m_{xi}, m_{yi}]^T$ and oriented as $\mathbf{v}_i = [\cos \alpha_i, \sin \alpha_i]^T$ measures the DOA $\vartheta_i = \theta_i + \alpha_i$ of an acoustic source located at \mathbf{s} in the far field.

combining a set of DOAs measured by a distributed network of N linear microphone arrays. With reference to Figure 1, the acoustic center of the i th array is at $\mathbf{m}_i = [m_{xi}, m_{yi}]^T$, $i = 1, \dots, N$, and the orientation of the line on which it lies is given by the unit vector $\mathbf{v}_i = [\cos \alpha_i, \sin \alpha_i]^T$. The vector \mathbf{v}_i forms the angle α_i with the x -axis. The DOA of the source, as measured with respect to the array line, is represented by the angle θ_i . The DOA is more conveniently represented in the global reference frame as $\vartheta_i = \theta_i + \alpha_i$.

Using (4), each DOA ϑ_i , $i = 1, \dots, N$, is turned into the corresponding acoustic ray $\mathbf{l}_i = [a_i, b_i, c_i]^T$. In absence of measurement noise, all these acoustic rays are bound to pass through the source position \mathbf{s} , meaning that

$$\begin{aligned} \mathbf{l}_1^T \mathbf{s} &= 0 \\ &\vdots \\ \mathbf{l}_N^T \mathbf{s} &= 0. \end{aligned} \quad (5)$$

In practical situations, DOAs are corrupted by measurement noise; thus the equalities in (5) do not hold. However, since the system is linear in the unknown variables (s_x, s_y) , the source position can be easily estimated as the solution of a linear least squares problem. To do so, we first rewrite the system as

$$\mathbf{M}\mathbf{s} = \mathbf{c}, \quad (6)$$

where

$$\mathbf{M} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \vdots & \vdots \\ a_N & b_N \end{bmatrix}, \quad \mathbf{s} = \begin{bmatrix} s_x \\ s_y \end{bmatrix},$$

$$\mathbf{c} = \begin{bmatrix} -c_1 \\ -c_2 \\ \vdots \\ -c_N \end{bmatrix}. \quad (7)$$

Then, the source position is estimated as

$$\hat{\mathbf{s}} = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{c}, \quad (8)$$

which corresponds to the solution of the linear least squares problem

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s}} \|\mathbf{M}\mathbf{s} - \mathbf{c}\|^2. \quad (9)$$

3. Theoretical Framework

Despite its simplicity, the solution provided in (8) is valid only for 2D configurations, that is, when the source and the microphone arrays lie on the same plane. For more general 3D scenarios, unfortunately, the generalization of the Ray Space is not straightforward. Indeed, representing acoustic rays in 3D requires the introduction of a five-dimensional projective space involving the use of Plücker's coordinates along with a number of nonlinear constraints [17], which would make the source localization problem intractable.

In this manuscript we follow an alternate route: rather than extending the idea of the Ray Space to a 3D setup, we focus on the reinterpretation of 3D source localization into an equivalent 2D space, by introducing the concept of equivalent microphone arrays.

3.1. DOAs in 3D Using Linear Arrays. A linear microphone array can be described by the parameters of a line passing through its center \mathbf{m}_i and directed as the unit vector \mathbf{v}_i . In a 3D setup, a total of 5 parameters are required. Specifically, three Cartesian coordinates define the center $\mathbf{m}_i = [m_{xi}, m_{yi}, m_{zi}]^T$, while the orientation is defined by the pair of angles α_i and β_i so that

$$\mathbf{v}_i = [\cos \alpha_i \cos \beta_i \quad \sin \alpha_i \cos \beta_i \quad \sin \beta_i]^T. \quad (10)$$

Using a linear array it is only possible to estimate a 2D DOA, which corresponds to the angle from which the source is observed by the array center, measured in the plane containing \mathbf{v}_i and the source position $\mathbf{s} = [s_x, s_y, s_z]^T$. More formally, the DOA as measured by the array is given by the angle formed between the vectors \mathbf{v}_i and $\mathbf{w}_i = \mathbf{s} - \mathbf{m}_i$, which can be computed as

$$\theta_i = \arccos \frac{\mathbf{v}_i^T \mathbf{w}_i}{\|\mathbf{v}_i\| \|\mathbf{w}_i\|} = \arccos \frac{\mathbf{v}_i^T \mathbf{w}_i}{\|\mathbf{w}_i\|}, \quad (11)$$

where we used the fact that $\|\mathbf{v}_i\| = 1$.

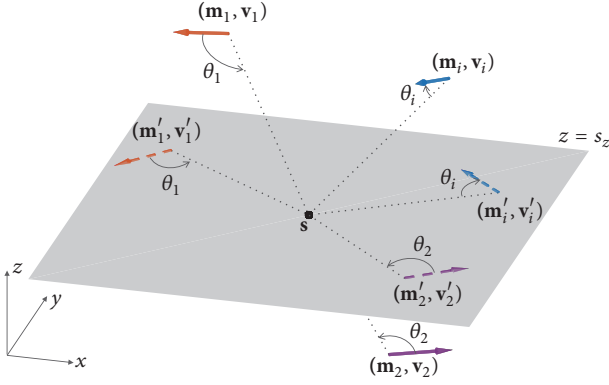


FIGURE 2: A set of *real* linear arrays (continuous colored arrows) are projected into the corresponding *equivalent* arrays (dashed colored arrows). Equivalent arrays all lie on the same plane, and they are disposed so that they look at the source from the same DOAs θ_i , $i = 1, \dots, N$, as measured by the real arrays.

3.2. Equivalent Microphone Arrays. Each DOA measurement refers to a different plane, according to the displacement and orientation of the arrays with respect to the source position. Nonetheless, we can interpret the available DOAs as being measured by a set of equivalent microphone arrays all located in the same plane, which contains also the source position \mathbf{s} . Analogously to the real ones, the equivalent arrays are described by the pairs $(\mathbf{m}'_i, \mathbf{v}'_i)$ indicating their centers and orientations, respectively, defined for $i = 1, \dots, N$ as

$$\begin{aligned} \mathbf{m}'_i &= \begin{bmatrix} m'_{xi} \\ m'_{yi} \\ m'_{zi} \end{bmatrix}, \\ \mathbf{v}'_i &= \begin{bmatrix} \cos \alpha'_i \cos \beta'_i \\ \sin \alpha'_i \cos \beta'_i \\ \sin \beta'_i \end{bmatrix}. \end{aligned} \quad (12)$$

This idea is illustrated in Figure 2, which highlights a set of real arrays disposed arbitrarily in the space and the corresponding equivalent arrays that all lie in the same plane. For clarity of visualization, we denoted each pair of real and equivalent arrays using the same color.

The final goal of this reinterpretation is to bring the 3D source localization problem to the one formulated in (9), which exploits the fact that all the measurements are referred to the same plane. For this reason, we must determine the unknown parameters \mathbf{m}'_i and \mathbf{v}'_i so that the equivalent arrays preserve the information related by their real counterparts. To do so, we notice that the only requirements are that all the equivalent arrays

- (i) must lie in a unique plane containing the source;
- (ii) must preserve the DOA, so that they observe the source from the same angles as the real arrays.

To satisfy the first constraint, without loss of generality, a convenient choice is to force all the arrays to lie on the plane $z = s_z$. Mathematically speaking, this corresponds to posing

$$m'_{zi} = s_z, \quad (13)$$

$$\beta'_i = 0, \quad (14)$$

for $i = 1, \dots, N$. As for the second requirement, formally we must require that the angle formed by the vectors \mathbf{v}'_i and $\mathbf{w}'_i = \mathbf{s} - \mathbf{m}'_i$ equals the DOA θ_i as defined in (11); that is,

$$\arccos \frac{\mathbf{v}'_i{}^T \mathbf{w}'_i}{\|\mathbf{w}'_i\|} = \arccos \frac{\mathbf{v}_i{}^T \mathbf{w}_i}{\|\mathbf{w}_i\|}, \quad (15)$$

which is equivalent to

$$\frac{\mathbf{v}'_i{}^T (\mathbf{s} - \mathbf{m}'_i)}{\|\mathbf{s} - \mathbf{m}'_i\|} = \frac{\mathbf{v}_i{}^T (\mathbf{s} - \mathbf{m}_i)}{\|\mathbf{s} - \mathbf{m}_i\|}. \quad (16)$$

We observe that the condition in (16) is met by an infinite number of equivalent arrays. This happens for two reasons, as exemplified in Figure 3. Indeed, on one hand, for a given choice of the angle α'_i , the DOA is preserved by all the parallel equivalent arrays directed as $\mathbf{v}'_i = [\cos \alpha'_i, \sin \alpha'_i, 0]^T$ and having centers on a common line, as in Figure 3(a). On the other hand, by keeping the distance ρ_i of the array from the source fixed, each rigid rotation of the array about \mathbf{s} leads to another possible equivalent array preserving the DOA, as shown in Figure 3(b). In order to make the solution unique, we impose that the transformation preserves the angle α_i (the orientation of the array on the horizontal plane) and the distance from the source. These additional constraints can be formally written as

$$\alpha'_i = \alpha_i, \quad (17)$$

$$\|\mathbf{s} - \mathbf{m}'_i\| = \|\mathbf{s} - \mathbf{m}_i\|, \quad (18)$$

where $i = 1, \dots, N$. Given the constraints in (13), (14), (17), and (18), what still remains undetermined are the positions of the equivalent arrays in the plane $z = s_z$, namely, the terms m'_{xi} and m'_{yi} . Since they depend on the unknown source location, we will derive m'_{xi} and m'_{yi} as functions of $\mathbf{s} = [s_x, s_y, s_z]^T$. To do so, we start by replacing (18) in (16) to obtain

$$\mathbf{v}'_i{}^T (\mathbf{s} - \mathbf{m}'_i) = \mathbf{v}_i{}^T (\mathbf{s} - \mathbf{m}_i). \quad (19)$$

By inserting (14), (13), and (17) into (19) and by making the terms explicit we obtain

$$\begin{aligned} & \begin{bmatrix} \cos \alpha'_i \\ \sin \alpha'_i \\ 0 \end{bmatrix}^T \begin{bmatrix} s_x - m'_{xi} \\ s_y - m'_{yi} \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} \cos \alpha_i \cos \beta_i \\ \sin \alpha_i \cos \beta_i \\ \sin \beta_i \end{bmatrix}^T \begin{bmatrix} s_x - m_{xi} \\ s_y - m_{yi} \\ s_z - m_{zi} \end{bmatrix}. \end{aligned} \quad (20)$$

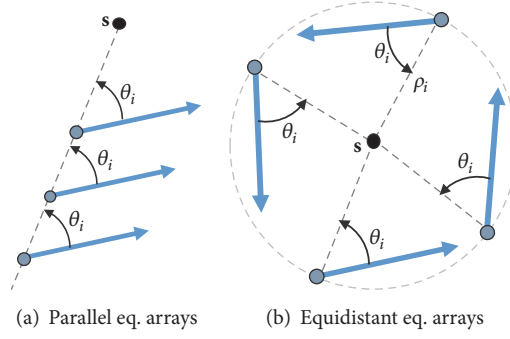


FIGURE 3: Examples of equivalent arrays all measuring the same DOA. In (a) the arrays are all parallel and with aligned centers. In (b) the arrays are equidistant from the source.

Moreover, if we take the square of (18) and we exploit the constraint in (13), we can write that

$$(s_x - m'_{xi})^2 + (s_y - m'_{yi})^2 = \|\mathbf{s} - \mathbf{m}_i\|^2. \quad (21)$$

Expressions in (20) and (21) form a second-order system in the unknown variables m'_{xi} and m'_{yi} . Rearranging the terms, this system can be rewritten in a more compact form as

$$\begin{aligned} P_i X_i + Q_i Y_i &= C_i, \\ X_i^2 + Y_i^2 &= D_i, \end{aligned} \quad (22)$$

with the following definitions:

$$\begin{aligned} X_i &= s_x - m'_{xi}, \\ Y_i &= s_y - m'_{yi}, \\ P_i &= \cos \alpha_i, \\ Q_i &= \sin \alpha_i, \\ C_i &= \mathbf{v}_i^T (\mathbf{s} - \mathbf{m}_i), \\ D_i &= \|\mathbf{s} - \mathbf{m}_i\|^2. \end{aligned} \quad (23)$$

Solving for X_i and Y_i we obtain

$$\begin{aligned} X_i &= \frac{C_i P_i \pm Q_i \sqrt{-C_i^2 + D_i (P_i^2 + Q_i^2)}}{P_i^2 + Q_i^2}, \\ Y_i &= \frac{C_i Q_i \mp P_i \sqrt{-C_i^2 + D_i (P_i^2 + Q_i^2)}}{P_i^2 + Q_i^2}. \end{aligned} \quad (24)$$

Using the definitions in (23) and noticing that $P_i^2 + Q_i^2 = 1$, we readily obtain the expressions for the equivalent array positions

$$\begin{aligned} m'_{xi} &= s_x - C_i P_i \pm Q_i \sqrt{D_i - C_i^2}, \\ m'_{yi} &= s_y - C_i Q_i \mp P_i \sqrt{D_i - C_i^2}. \end{aligned} \quad (25)$$

Note that, as expected, the second-order system (22) leads to two valid solutions. They represent two equivalent arrays,

mutually mirrored about the source. Consequently, we can safely select one of those randomly, without loss of information. For instance, one may select the equivalent array located closest to the corresponding real one.

4. Source Localization

In this section we take advantage of the planar geometry induced by the transformation of the real into the equivalent arrays for the purposes of source localization. We first reformulate the source localization problem as a nonlinear least squares minimization problem, introducing the equivalent array positions in the cost function in (9). Then, we describe the iterative minimization process leading to the estimation of the source location. Finally, we describe how the estimated location can be refined by detecting and removing potential outliers from the set of measurements.

4.1. Cost Function Definition. The introduction of the equivalent arrays makes it possible to exploit the 2D Ray Space localization discussed in Section 2.3. Let us consider a network of N linear arrays arbitrarily deployed in the 3D space, whose positions \mathbf{m}_i and orientations \mathbf{v}_i are assumed to be known. For a candidate source position \mathbf{s} , the positions of the equivalent array centers can be computed using (25), here denoted as $m'_{xi}(\mathbf{s})$ and $m'_{yi}(\mathbf{s})$ to stress the dependency on \mathbf{s} . With α_i being the direction angle of the equivalent array and θ_i the DOA in the array reference system, the DOA expressed in the global reference system (i.e., measured in the plane $z = s_z$ with respect to the x axis) is therefore $\vartheta_i = \theta_i + \alpha_i$. Consequently, the acoustic ray associated with the DOA measured by the i th equivalent array turns to be $\mathbf{l}_i(\mathbf{s}) = [a_i, b_i, c_i(\mathbf{s})]^T$, whose parameters can be computed using (4), obtaining

$$\begin{aligned} a_i &= \sin \vartheta_i, \\ b_i &= -\cos \vartheta_i, \end{aligned} \quad (26)$$

$$c_i(\mathbf{s}) = m'_{xi}(\mathbf{s}) \sin \vartheta_i - m'_{yi}(\mathbf{s}) \cos \vartheta_i.$$

Differently from the pure 2D case discussed in Section 2.3, the parameter $c_i(\mathbf{s})$ of the acoustic ray is a nonlinear function of the source position \mathbf{s} . For this reason, the combination of N equations in the form $\mathbf{l}_i^T(\mathbf{s})\mathbf{s} = 0$ does not lead anymore to the

definition of a linear system. Nevertheless, we can still formulate the source localization problem as in (9), where now the vector \mathbf{c} depends nonlinearly on the source location \mathbf{s} . More formally, the optimization problem can be rewritten as the minimization of the cost function $J(\mathbf{s}) = \mathbf{e}^T(\mathbf{s})\mathbf{e}(\mathbf{s})$; that is,

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s}} J(\mathbf{s}) = \arg \min_{\mathbf{s}} \mathbf{e}^T(\mathbf{s})\mathbf{e}(\mathbf{s}), \quad (27)$$

where $\mathbf{e}(\mathbf{s}) = \mathbf{M}\mathbf{s} - \mathbf{c}(\mathbf{s})$ is the error vector (i.e., the localization residuals) and $\mathbf{c}(\mathbf{s}) = [-c_1(\mathbf{s}), \dots, -c_N(\mathbf{s})]^T$.

4.2. Minimization. The minimization of the cost function can be efficiently addressed in an iterative fashion, by locally approximating $J(\mathbf{s})$ at each iteration by means of a Taylor series expansion. Following the same approach as in [19], we compute the first-order expansion of the localization residual about an initial guess position $\mathbf{s}^{(0)} = [s_x^{(0)}, s_y^{(0)}, s_z^{(0)}]^T$ as

$$\mathbf{e}(\mathbf{s}) \simeq \mathbf{e}(\mathbf{s})|_{\mathbf{s}^{(0)}} + \nabla_{\mathbf{e}}|_{\mathbf{s}^{(0)}} \cdot (\mathbf{s} - \mathbf{s}^{(0)}). \quad (28)$$

The Jacobian matrix $\nabla_{\mathbf{e}}$ is given by

$$\nabla_{\mathbf{e}} = [\nabla e_1(\mathbf{s}) \ \dots \ \nabla e_N(\mathbf{s})]^T, \quad (29)$$

where

$$\nabla e_i(\mathbf{s}) = \left[\frac{\partial e_i(\mathbf{s})}{\partial s_x} \ \frac{\partial e_i(\mathbf{s})}{\partial s_y} \ \frac{\partial e_i(\mathbf{s})}{\partial s_z} \right]^T, \quad (30)$$

$$e_i(\mathbf{s}) = a_i s_x + b_i s_y - c_i(\mathbf{s}), \quad i = 1, \dots, N.$$

According to the problem in (27), the update equation of the iterative update is given by [19]

$$\mathbf{s}^{(t+1)} = \mathbf{s}^{(t)} - \nabla_{\mathbf{e}}^\dagger|_{\mathbf{s}^{(t)}} \cdot \mathbf{e}(\mathbf{s})|_{\mathbf{s}^{(t)}}, \quad (31)$$

where t is the iteration number and

$$\nabla_{\mathbf{e}}^\dagger|_{\mathbf{s}^{(t)}} = \left(\nabla_{\mathbf{e}}^T|_{\mathbf{s}^{(t)}} \cdot \nabla_{\mathbf{e}}|_{\mathbf{s}^{(t)}} \right)^{-1} \cdot \nabla_{\mathbf{e}}|_{\mathbf{s}^{(t)}}. \quad (32)$$

The source location is estimated as

$$\hat{\mathbf{s}} = \mathbf{s}^{(t+1)}, \quad (33)$$

after stopping the iteration when $|\mathbf{s}^{(t+1)} - \mathbf{s}^{(t)}|$ is smaller than a prescribed threshold.

4.3. Localization Refinement. In adverse acoustic conditions, the DOA set may contain some outliers, typically related to the bias introduced by reverberation in standard DOA estimation algorithms [20]. A simple and effective method for detecting outliers is to check the magnitude of localization residuals $|e_i(\hat{\mathbf{s}})|$, $i = 1, \dots, N$. Inspired by the work in [21] related to time differences of arrival-based source localization, we compute the standard deviation $\sigma_{\mathcal{E}}$ of the residual set $\mathcal{E} = \{e_1(\hat{\mathbf{s}}), \dots, e_N(\hat{\mathbf{s}})\}$. If $\sigma_{\mathcal{E}}$ exceeds a prescribed threshold $T_{\mathcal{E}}$, the DOA set potentially contains outliers. In this case, we mark as an outlier the DOA θ_i associated with the maximum localization residual, that is, so that

$$\bar{i} = \arg \max_i |e_i(\hat{\mathbf{s}})|. \quad (34)$$

This outlier DOA $\theta_{\bar{i}}$ is thus removed from the measurement set, and the associated residual $e_{\bar{i}}(\hat{\mathbf{s}})$ is removed from \mathcal{E} .

TABLE 1: Array positions and orientations.

Array	\mathbf{m}_i	α_i	β_i
$i = 1$	$[0.5 \ 0.5 \ 0.5]^T$	-135°	45°
$i = 2$	$[3.5 \ 0.5 \ 0.5]^T$	45°	0°
$i = 3$	$[3.5 \ 2.5 \ 0.5]^T$	45°	0°
$i = 4$	$[0.5 \ 2.5 \ 0.5]^T$	225°	0°
$i = 5$	$[0.5 \ 0.5 \ 2.5]^T$	-45°	0°
$i = 6$	$[3.5 \ 0.5 \ 2.5]^T$	135°	45°
$i = 7$	$[3.5 \ 2.5 \ 2.5]^T$	135°	0°
$i = 8$	$[0.5 \ 2.5 \ 2.5]^T$	-45°	45°

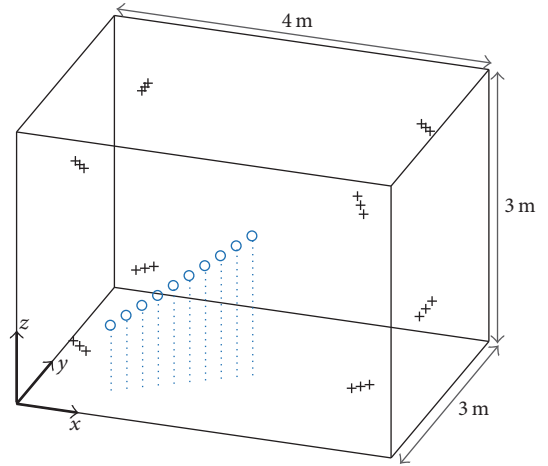


FIGURE 4: Simulation setup: $N = 8$ microphone arrays are disposed near to the vertices of a rectangular room. Blue circles denote the tested source positions.

The standard deviation $\sigma_{\mathcal{E}}$ is then recomputed, and the procedure is iterated until $\sigma_{\mathcal{E}} \leq T_{\mathcal{E}}$ or a maximum number of measurements are removed. At this point, the source location is refined by running a second time the algorithm presented in Section 4.2, recomputing the cost function with the outlier-free measurement set.

As discussed in [21], it is worth noticing that this refinement step has a negligible computational cost, as the residual terms are directly available after source localization. Moreover, the iterative minimization procedure is repeated at most one time. Note also that the localization refinement is accomplished by the central node and does not require additional data exchange with the sensors.

5. Validation

We now present the results of a set of simulations aimed at verifying the validity of the proposed method. With reference to the setup shown in Figure 4, we tested the localization algorithm in a rectangular room of size $4 \times 3 \times 3 \text{ m}^3$ by means of $N = 8$ linear arrays located close to the vertices of the room. Each sensor is equipped with 3 microphones, located on a line segment and uniformly spaced by $r = 10 \text{ cm}$. The positions \mathbf{m}_i and orientations (α_i, β_i) of the sensors are summarized in Table 1. We considered 10 test source

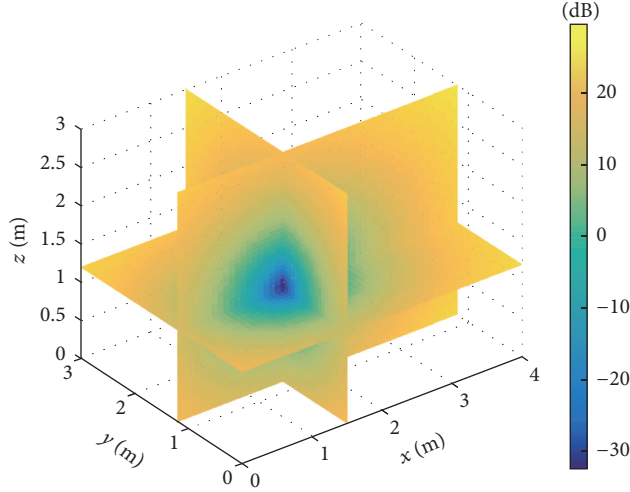


FIGURE 5: Planar sections of the 3D cost function $J(\mathbf{s})$ for a selected test source position at $\mathbf{s} = [1.5 \text{ m}, 1.2 \text{ m}, 1.2 \text{ m}]^T$. Three planar sections are displayed, in correspondence of the planes $x = 1.5 \text{ m}$, $y = 1.2 \text{ m}$, and $z = 1.2 \text{ m}$, respectively.

positions, located on the line connecting the center of the room and the center of the first sensor.

5.1. Test 1: Robustness against Additive Noise on DOAs. In a first stage, we considered a completely anechoic scenario, in which we simulated the measurement of noisy DOAs. This allowed us to test the statistical behavior of the proposed localization cost function. More specifically, we corrupted the nominal DOAs $\bar{\theta}_i$, $i = 1, \dots, N$, observed by each array with additive noise ϵ_i , which is a realization of zero-mean white noise drawn from a normal distribution with standard deviation σ_θ . Note that, in this case, the measurements are by definition free of outliers, since the deviation from the nominal DOA values is only due to the additive error model [22].

In order to enforce the statistical validity of the analysis of the cost function, we took care of two aspects. First, we bypassed the source localization refinement step by setting $T_{\mathcal{G}} = \infty$. This avoided the risk of incidentally removing DOAs from the measurement set, which is explicitly free of outliers. Second, we verified (at least empirically) that the iterative minimization procedure converges to the global optimum. Even though the cost function $J(\mathbf{s})$ is not convex, for the considered setup we observed that it presents a very smooth behavior, along with a wide basin of attraction around its global minimum. This turns to be true independently from the tested source positions. These facts ensure a rapid convergence of the iterative algorithms towards the solution, without significant risks of getting trapped into local optima. An illustrative example is offered in Figure 5, which shows three planar sections of $J(\mathbf{s})$ corresponding to a randomly picked realization of noisy DOAs, for a selected source location. In particular, the source is located at $\mathbf{s} = [1.5 \text{ m}, 1.2 \text{ m}, 1.2 \text{ m}]^T$ and the cost function is shown in three orthogonal planar sections passing through \mathbf{s} ; that is, at $x = 1.5 \text{ m}$, $y = 1.2 \text{ m}$, and $z = 1.2 \text{ m}$.

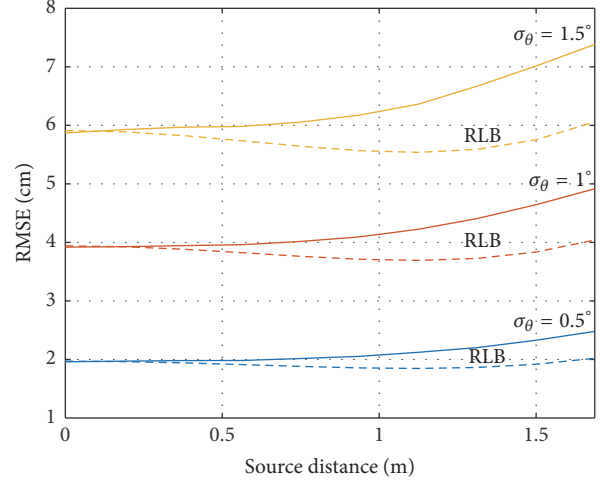


FIGURE 6: RMSE as a function of the source position, expressed in terms of its distance from the room center.

For the statistical analysis, we considered different noise standard deviations in the range $\sigma_\theta \in [0.25^\circ, 3^\circ]$. For each value of σ_θ and for each tested source position we run 2500 Monte-Carlo repetitions. The localization accuracy was measured in terms of root means squared error, computed as

$$\text{RMSE}(\mathbf{s}) = \sqrt{\frac{1}{L} \sum_{l=1}^L \|\hat{\mathbf{s}}_l - \mathbf{s}\|^2}, \quad (35)$$

where \mathbf{s} is the true source location and $\hat{\mathbf{s}}_l$ denotes its estimate at the l th Monte-Carlo run and L is the number of realizations.

As a matter of comparison, we computed the lower bound for the RMSE implied by the Cramer-Rao Lower Bound (CRLB). More specifically, given the CRLB in terms of the localization variances $\sigma_{s_x}^2(\mathbf{s})$, $\sigma_{s_y}^2(\mathbf{s})$, and $\sigma_{s_z}^2(\mathbf{s})$, the RMSE Lower Bound (RLB) was computed as

$$\text{RLB}(\mathbf{s}) = \sqrt{\sigma_{s_x}^2(\mathbf{s}) + \sigma_{s_y}^2(\mathbf{s}) + \sigma_{s_z}^2(\mathbf{s})}. \quad (36)$$

Details about the computation of the CRLB are given in Appendix.

In Figure 6 we report the resulting RMSE as a function of the distance of the source from the center of the room, for three selected noise standard deviations. The continuous lines refer to the RMSE, while the dashed lines denote the related RLB. We observe that RMSE ranges from a minimum of about 2 cm for $\sigma_\theta = 0.5^\circ$ to a maximum of about 7.5 cm for $\sigma_\theta = 1.5^\circ$. The RMSE tends to increase as the source position approaches the center of the first array. Moreover, we notice that the RMSE approaches the RLB for source positions close to the room center, while the accuracy slightly degrades for farther sources. Nevertheless, the maximum gap between the RMSE and the RLB maintains reasonable, being slightly above 1 cm in the worst case.

Inspired by this observation, we evaluated the RMSE as a function of the standard deviation of the injected noise for

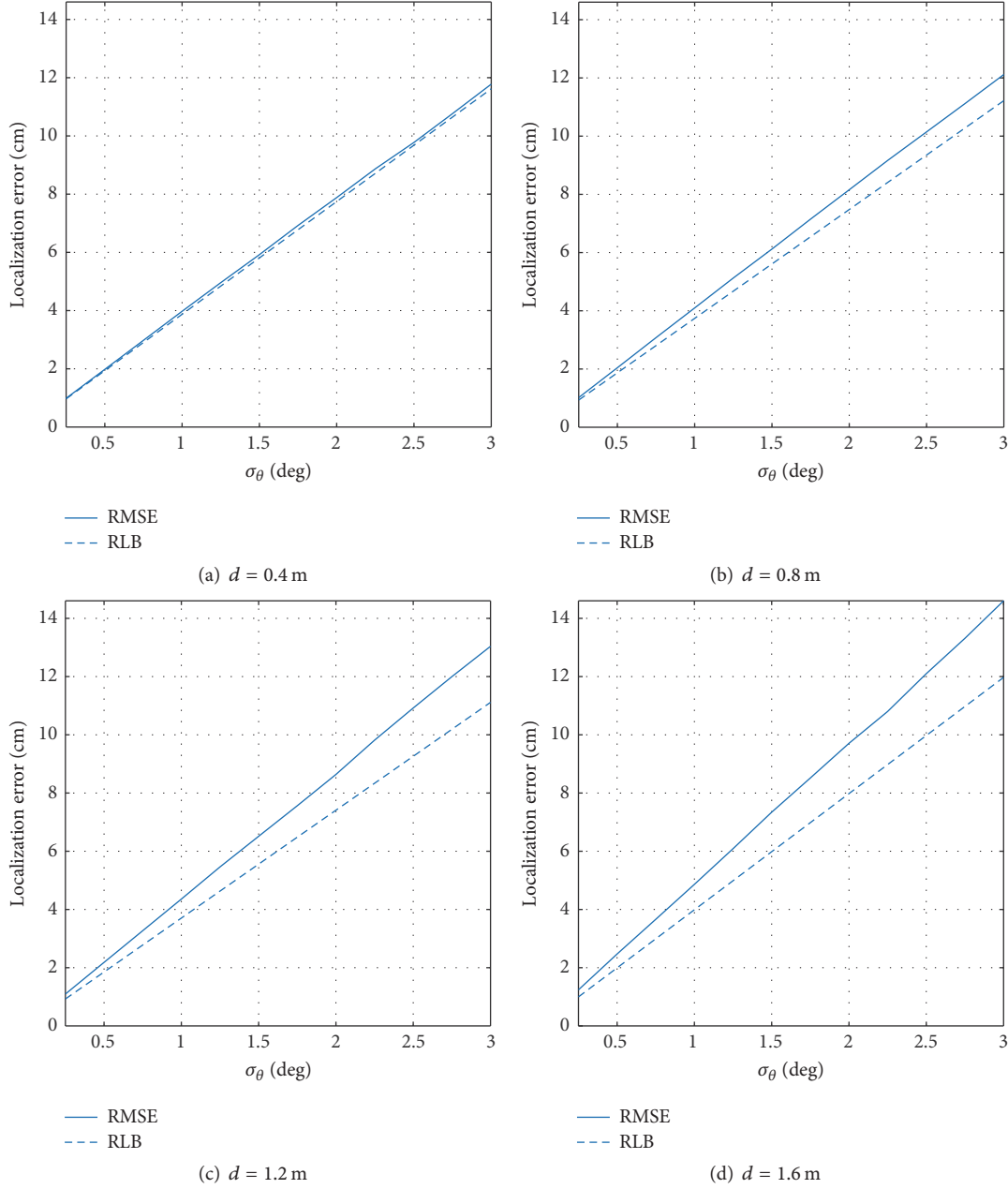


FIGURE 7: RMSE as a function of the noise standard deviation, for source located at different distances d from the room center.

4 selected source positions located at different distances d from the room center. Results are reported in Figure 7. We first notice that the RMSE grows almost linearly with respect to the standard deviation σ_θ . Moreover, once again we notice that the gap between the RMSE and the RLB tends to increase when the distance d of the source from the room center increases. In the worst case (i.e., when $\sigma_\theta = 3^\circ$ and $d = 1.6$ m) this gap is about 2 cm; thus we can conclude that the proposed method is sufficiently robust against Gaussian additive noise on DOAs.

5.2. Test 2: Robustness against Reverberation and Noise. To mimic a realistic scenario, we simulated microphone signal

acquisitions in a reverberant room by means of the image source method [23]. In particular, we adopted the implementation provided in the RIR toolbox [24] to generate the room impulse responses from the source position to all the microphones. For this test we considered 50 test source positions randomly selected within the rectangular volume formed by the 8 array centers. The acoustic source was modeled to be omnidirectional and emitted 30 s of phonetically rich female speech. The microphone signals were obtained by convolving the source signal with the related RIRs and finally corrupted with zero-mean Gaussian noise with standard deviation tuned to simulate a prescribed signal-to-noise ratio (SNR). The signals were divided into

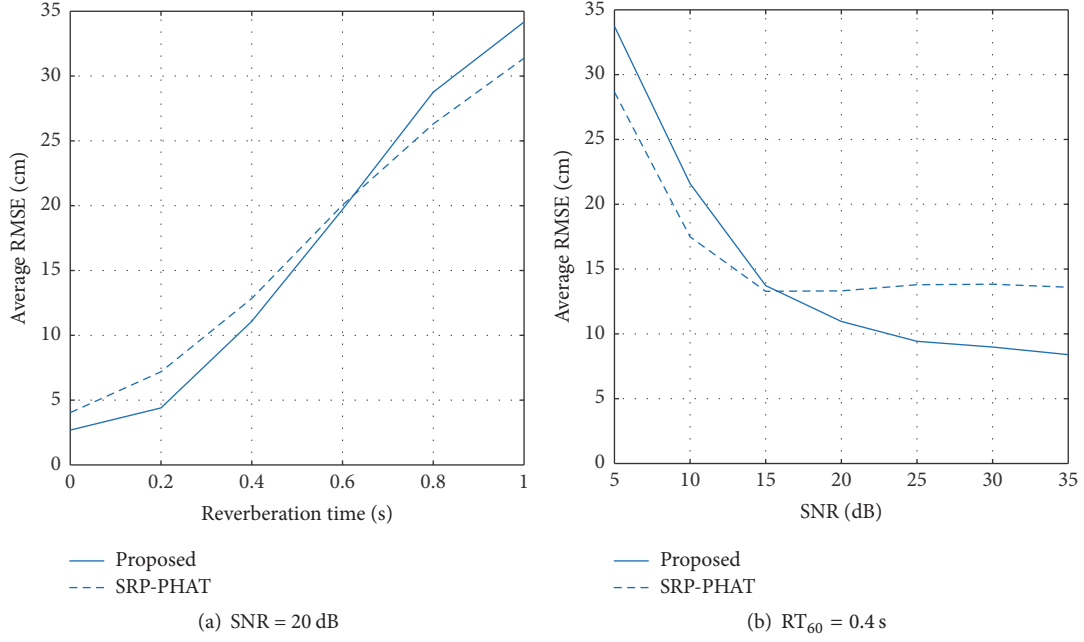


FIGURE 8: Localization accuracy at different reverberation and SNR levels.

frames of length 2048 samples at a sampling frequency $f_s = 16$ kHz. DOA estimation was performed by using a beamforming technique. Specifically, for each array, the received microphone signals were transformed into the time-frequency domain via short-time Fourier transform (STFT), considering a Hamming analysis windows of length 256 samples, with 50% of overlap, leading to a total of $W = 15$ time windows per frame. Let $\mathbf{x}_i(n, k)$ be the STFT representation of a generic signal frame acquired by the i th array, where n is the index of the time window and k is the bin associated with the frequency ω_k . At each frequency bin, we computed the Minimum Variance Distortionless Response (MVDR) pseudospectra [14] as

$$p_i(\phi, k) = \frac{1}{\mathbf{a}^T(\phi, k) \mathbf{G}^{-1}(k) \mathbf{a}(\phi, k)}, \quad (37)$$

where $\mathbf{a}(\phi, k) = [e^{j\text{rsin}(\phi)\omega_k}, 0, e^{-j\text{rsin}(\phi)\omega_k}]^T$ is the far-field steering vector for the used arrays, each composed of three microphones, and $\mathbf{G}(k)$ is the sample estimate of the array covariance matrix, obtained as

$$\mathbf{G}(k) = \frac{1}{W} \sum_{n=1}^W \mathbf{x}(n, k) \mathbf{x}(n, k)^H. \quad (38)$$

The DOA at the i th array is finally estimated as

$$\theta_i = \arg \max_{\phi} P_i(\phi), \quad (39)$$

where $P_i(\phi)$ is the geometric mean of the pseudospectra $p_i(\phi, k)$ with frequency bins k in the range [250 Hz, 8 kHz]. The collected DOAs θ_i , $i = 1, \dots, N$, were then used to estimate source position with the proposed method. For this test, we enabled the localization refinement step by selecting

$T_{\mathcal{E}} = 10$ cm, while ensuring that a maximum of 2 potential outliers are removed from the DOA set.

As a matter of comparison, we estimated the source position also using the popular steered-response power with phase transform (SRP-PHAT) method, which is known to be very robust against both noise and reverberation. In particular, we used the Stochastic Region Contraction variant of SRP (SRC-PHAT) [25], which is best suited for 3D scenarios. To enable a fair comparison, the generalized cross correlations (GCC-PHATs) were computed only at microphone pairs belonging to the same linear array, for a total of $3 \times N$ GCC-PHAT computations. Thinking in terms of distributed localization, these GCC-PHATs functions are transferred to the central node that finally combines them to produce the source position estimate.

The results are reported in Figures 8(a) and 8(b), showing the RMSE achieved by the proposed method and SRP-PHAT, averaged among all the tested source positions and all the analyzed audio frames. In particular, Figure 8(a) reports the localization accuracy at different levels of reverberation, expressed in terms of the reverberation time RT_{60} [26], while keeping the SNR fixed to 20 dB. Conversely, Figure 8(b) shows the localization accuracy at different SNR levels, while the reverberation is kept fixed to $RT_{60} \leq 0.4$ s. It can be noticed that the proposed technique achieves better scores than SRP-PHAT up to a moderate amount of reverberation $RT_{60} \leq 0.6$ s and above an acceptable SNR level (i.e., greater than 15 dB). Out of these ranges, SRP-PHAT is slightly more robust against reverberation and noise.

5.3. Discussion. Even though SRP-PHAT tends to be more robust in particularly adverse acoustic environments, this comes at the expense of both higher computational complexity and higher communication bandwidth requirements.

Let us consider again the distributed source localization scenario, in which source localization is demanded to a central node that gathers measurements coming from the distributed arrays. For both the considered methods, the computational power required at the sensors is comparable (beamforming for the proposed technique, computation of generalized cross correlations for SRC-PHAT). However, as far as the bandwidth is concerned, the proposed method only has to transmit a single real value to the central node, that is, the DOA. Conversely, each sensor must transmit the entire GCC-PHATs to the central node when localization is tackled via SRC-PHAT. Moreover, the computational burden at the central node is much higher for the SRC-PHAT method, being based on the minimization of a very irregular 3D cost function which requires thousands of evaluations of the SRC functional [21]. On the other hand, being based on a very smooth cost function, the proposed method requires a few iterations to converge to the optimum solution. As an example, for the considered simulation scenario, we measured the execution times at the central node, which highlighted that, on the average, the proposed method (implemented in Matlab®) is more than 10 times faster than the SRC-PHAT Matlab implementation (<https://it.mathworks.com/matlabcentral/fileexchange/24352-acoustic-source-localization-using-srp-phat>) released by the authors of [25]. More specifically, algorithms at the central node were run on a laptop equipped with a 2.9 GHz Intel Core i7 processor and 8 GB of RAM. The average time to process a 128 ms-long audio frame was 3.26 ms for the proposed method and 39.89 ms for the SRC-PHAT algorithm. It is worth noticing that this result is only qualitative, as the algorithm implementation is not optimized. Nevertheless, this suggests that the proposed method may represent a valid alternative for low-bandwidth and low-power application scenarios.

6. Conclusions

In this manuscript we presented a technique for 3D source localization based on the combination of two-dimensional DOAs. The proposed methodology works in two steps: first real arrays are transformed into equivalent arrays so that they are coplanar with the source. In the second step, the source is localized on the plane on which equivalent arrays and source lie. The main benefit of the proposed technique lies in the hardware and computational costs. Indeed, the measurement of 2D DOAs can be accomplished using linear arrays and involves a one-dimensional grid search. Conversely, state-of-the-art techniques for 3D source localization are based on 3D DOAs, which require more complex microphone arrangements as well as a higher computational complexity. The proposed technique is validated with an extensive simulation campaign and compared with Stochastic Region Contraction variant of SRP (SRC-PHAT). Results demonstrate that in moderately reverberant environments and in the presence of moderate noise, the proposed technique achieves results comparable with SRC-PHAT, which requires a much greater effort in terms of bandwidth and computational cost.

We are currently working on the extension of the proposed algorithm to multiple sources. Another important

issue that deserves to be investigated is that of the sensitivity of the localization accuracy towards noise on the estimated DOAs and how the configuration of the arrays impacts on the sensitivity.

Appendix

Cramer-Rao Lower Bound

The Cramer-Rao Lower Bound (CRLB) expresses the theoretical limit for the variance of an unbiased estimator. We consider here the problem of estimating the source position \mathbf{s} from DOA measurements at N linear arrays. To compute the CRLB, we follow the same rationale of [1] for localization based on time differences of arrival. We start modeling the DOA measurements as

$$\theta_i = g_i(\mathbf{s}) + \epsilon_i, \quad g_i(\mathbf{s}) = \arccos \frac{\mathbf{v}_i^T (\mathbf{s} - \mathbf{m}_i)}{\|\mathbf{s} - \mathbf{m}_i\|}, \quad (\text{A.1})$$

where $g_i(\mathbf{s})$ is the noiseless DOA and ϵ_i is the measurement error for the i th array. Expressing the measurement model in a vector form, we have

$$\boldsymbol{\theta} = \mathbf{g}(\mathbf{s}) + \boldsymbol{\epsilon}, \quad (\text{A.2})$$

where $\boldsymbol{\theta}(\mathbf{s}) = [\theta_1, \dots, \theta_N]^T$, $\mathbf{g}(\mathbf{s}) = [g_1(\mathbf{s}), \dots, g_N(\mathbf{s})]^T$, and $\boldsymbol{\epsilon} = [\epsilon_1, \dots, \epsilon_N]^T$.

The CRLB is given by the elements of the inverse of the Fisher information matrix, which is in general computed as

$$\mathbf{F}(\mathbf{s}) = -E \left\{ \nabla^2 [\ln p(\boldsymbol{\theta} | \mathbf{s})] \right\}, \quad (\text{A.3})$$

where $\nabla^2 [\ln p(\boldsymbol{\theta} | \mathbf{s})]$ is the Hessian matrix of the function $\ln p(\boldsymbol{\theta} | \mathbf{s})$ and $p(\boldsymbol{\theta} | \mathbf{s})$ is the probability density function (PDF) of measuring $\boldsymbol{\theta}$ given the source position \mathbf{s} . More specifically, the CRLBs of the variance of the estimation of s_x , s_y , and s_z are obtained as

$$\left[\sigma_{s_x}(\mathbf{s}), \sigma_{s_y}(\mathbf{s}), \sigma_{s_z}(\mathbf{s}) \right]^T = \text{diag} \left[\mathbf{F}^{-1}(\mathbf{s}) \right]. \quad (\text{A.4})$$

Assuming that the noise ϵ_i corrupting the DOAs is jointly Gaussian distributed, and independent of both the observed DOAs and the source position, the PDF reads as

$$p(\boldsymbol{\theta} | \mathbf{s}) = \frac{e^{-(1/2)[\boldsymbol{\theta} - \mathbf{g}(\mathbf{s})]^T \boldsymbol{\Sigma}_\epsilon^{-1} [\boldsymbol{\theta} - \mathbf{g}(\mathbf{s})]}}{\sqrt{(2\pi)^N \det(\boldsymbol{\Sigma}_\epsilon)}}, \quad (\text{A.5})$$

where $\boldsymbol{\Sigma}_\epsilon$ is the covariance matrix of the measurement error $\boldsymbol{\epsilon}$. In this case, the Fisher information matrix simplifies to

$$\mathbf{F}(\mathbf{s}) = \left[\frac{\partial \mathbf{g}(\mathbf{s})}{\partial \mathbf{s}} \right]^T \boldsymbol{\Sigma}_\epsilon^{-1} \left[\frac{\partial \mathbf{g}(\mathbf{s})}{\partial \mathbf{s}} \right], \quad (\text{A.6})$$

where the Jacobian matrix is given by

$$\frac{\partial \mathbf{g}(\mathbf{s})}{\partial \mathbf{s}} = [\nabla g_1(\mathbf{s}) \quad \nabla g_2(\mathbf{s}) \quad \dots \quad \nabla g_N(\mathbf{s})]^T. \quad (\text{A.7})$$

The gradient vectors $\nabla g_i(\mathbf{s})$ can be computed in closed form as

$$\nabla g_i(\mathbf{s}) = \frac{(\mathbf{s} - \mathbf{m}_i) [\mathbf{v}_i^T (\mathbf{s} - \mathbf{m}_i)] / \|\mathbf{s} - \mathbf{m}_i\|^3 - \mathbf{v}_i / \|\mathbf{s} - \mathbf{m}_i\|}{\sqrt{1 - \mathbf{v}_i^T (\mathbf{s} - \mathbf{m}_i) / \|\mathbf{s} - \mathbf{m}_i\|}} \quad (\text{A.8})$$

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work is supported by “CARIPLO Foundation” and “Regione Lombardia” under Grant no. 2016-0917 (SPIN project).

References

- [1] Y. Huang, *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, chapter 8: Source Localization, Springer US, 2004.
- [2] B. Dahlan, W. Mansoor, M. Abbasi, and P. Honarbakhsh, “Sound source localization for automatic camera steering,” in *Proceedings of the 7th International Conference on Networked Computing and Advanced Information Management (NCM '11)*, pp. 20–25, June 2011.
- [3] Y. Huang, J. Benesty, and G. W. Elko, “Passive acoustic source localization for video camera steering,” in *Proceedings of the 25th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '00)*, vol. 2, pp. II909–II912, June 2000.
- [4] G. Valenzise, L. Gerosa, M. Tagliasacchi, F. Antonacci, and A. Sarti, “Scream and gunshot detection and localization for audio-surveillance systems,” in *Proceedings of the 2007 IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS '07)*, pp. 21–26, September 2007.
- [5] S. Argentieri and P. Danès, “Broadband variations of the MUSIC high-resolution method for sound source localization in robotics,” in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '07)*, pp. 2009–2014, October 2007.
- [6] R. Roy and T. Kailath, “ESPRIT-estimation of signal parameters via rotational invariance techniques,” *IEEE Transactions on Signal Processing*, vol. 37, no. 7, pp. 984–995, July 1989.
- [7] A. Griffin, A. Alexandridis, D. Pavlidis, Y. Mastorakis, and A. Mouchtaris, “Localizing multiple audio sources in a wireless acoustic sensor network,” *Signal Processing*, vol. 107, pp. 54–67, 2015.
- [8] A. Alexandridis and A. Mouchtaris, “Multiple sound source location estimation and counting in a wireless acoustic sensor network,” in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA '15)*, pp. 1–5, October 2015.
- [9] A. Alexandridis, G. Borboudakis, and A. Mouchtaris, “Addressing the data-association problem for multiple sound source localization using DOA estimates,” in *Proceedings of the 23rd European Signal Processing Conference (EUSIPCO '15)*, pp. 1551–1555, August 2015.
- [10] K. Doğançay, “Bearings-only target localization using total least squares,” *Signal Processing*, vol. 85, no. 9, pp. 1695–1710, 2005.
- [11] A. N. Bishop, B. D. O. Anderson, B. Fidan, P. N. Pathirana, and G. Mao, “Bearing-only localization using geometrically constrained optimization,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 1, pp. 308–320, 2009.
- [12] L. M. Kaplan, Q. Le, and N. Molnar, “Maximum likelihood methods for bearings-only target localization,” in *Proceedings of the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 3001–3004, 2001.
- [13] X. Pan, H. Wang, F. Wang, and C. Song, “Multiple spherical arrays design for acoustic source localization,” in *Proceedings of the 6th Conference of the Sensor Signal Processing for Defence (SSPD '16)*, pp. 1–5, September 2016.
- [14] P. Stoica and R. L. Moses, “Introduction to spectral analysis,” *Prentice hall Upper Saddle River*, vol. 1, 1997.
- [15] M. Brandstein and D. Ward, “Microphone arrays: signal processing techniques and applications,” in *Digital Signal Processing*, Springer-Verlag, 2001.
- [16] R. Levorato and E. Pagello, “DOA acoustic source localization in mobile robot sensor networks,” in *Proceedings of the 9th IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC '15)*, pp. 71–76, April 2015.
- [17] D. Markovic, F. Antonacci, A. Sarti, and S. Tubaro, “3D beam tracing based on visibility lookup for interactive acoustic modeling,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 10, pp. 2262–2274, 2016.
- [18] D. Marković, A. Canclini, F. Antonacci, A. Sarti, and S. Tubaro, “Visibility-based beam tracing for soundfield rendering,” in *Proceedings of the 2010 IEEE International Workshop on Multimedia Signal Processing (MMSP '10)*, pp. 40–45, October 2010.
- [19] M. Compagnoni, P. Bestagini, F. Antonacci, A. Sarti, and S. Tubaro, “Localization of acoustic sources through the fitting of propagation cones using multiple independent arrays,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 7, pp. 1964–1975, 2012.
- [20] F. Jacob and R. Haeb-Umbach, “On the bias of direction of arrival estimation using linear microphone arrays,” in *Proceedings of the Speech Communication; 12. ITG Symposium*, pp. 1–5, October 2016.
- [21] A. Canclini, P. Bestagini, F. Antonacci, M. Compagnoni, A. Sarti, and S. Tubaro, “A robust and low-complexity source localization algorithm for asynchronous distributed microphone networks,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 23, no. 10, pp. 1563–1575, 2015.
- [22] M. Compagnoni, A. Pini, A. Canclini et al., “A geometrical-statistical approach to outlier removal for TDOA measurements,” *IEEE Transactions on Signal Processing*, vol. 65, no. 15, pp. 3960–3975, 2017.
- [23] J. B. Allen and D. A. Berkley, “Image method for efficiently simulating small-room acoustics,” *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.
- [24] “Room impulse response generator,” <https://github.com/ehabets/RIR-generator>.
- [25] H. Do, H. F. Silverman, and Y. Yu, “A real-time SRP-PHAT source location implementation using stochastic region contraction (SRC) on a large-aperture microphone array,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, vol. 1, pp. 1–121–1–124, April 2007.
- [26] H. Kuttruff, *Room Acoustics*, CRC Press, 2016.

